



22	Audio output part
23	Video output part
25, 26	D/A converter
101	Recording medium
102	Memory bus controller
104	Recording medium I/F

### Claims

1. A device for processing image data and audio data, characterized in that it is equipped with a recording medium that records image data and audio data by correlating them,

a specification means that specifies audio data stored in the aforementioned recording medium,

a control means that retrieves image data that are correlated with the aforementioned specified audio data in the form of file information, and

a reading means that reads the aforementioned retrieved image data and reproduces them.

2. The device for processing image data and audio data described in Claim 1, characterized in that it is further equipped with a character recognition means that performs character recognition of the aforementioned image data, whereby it performs registration processing so as to record the result of recognition by said character recognition means in the form of a text file by correlating it with the aforementioned respective image and audio data while allowing it to be retrieved.

3. The device for processing image data and audio data described in Claim 1, characterized in that it is further equipped with a speech recognition means that recognizes the aforementioned audio data in the form of characters, whereby it performs registration processing so as to record the result of recognition by said speech recognition means in the form of a text file by correlating it with the aforementioned respective image and audio data while allowing it to be retrieved.

4. The device for processing image data and audio data described in Claim 1, characterized in that it is further equipped with a display means that displays the aforementioned respective data stored in the aforementioned recording medium in the form of file information, whereby mutually correlated information among the aforementioned respective files is edited, and the results of retrieval of the aforementioned respective data based on said correlated information are displayed.

Detailed explanation of the invention

[0001]

Industrial application field

The present invention pertains to a device that processes image and audio so as to record and play them.

[0002]

Prior art

A still video format has been known as a standard for processing and recording image and audio data. In the case of said conventional still video format, image data and audio data are FM-recorded separately on different tracks. In addition, the audio track is provided with a field that is used to record the track number of a correlated image track in its control code and is used to record a note (identification information) regarding a specific image.

[0003]

Problems to be solved by the invention

However, the aforementioned conventional digital electronic camera has the following problems.

① That is, with the still video format, although an image track to be looked up from an audio track could be specified, the correlated audio track could not be looked up from the image track. Thus, there was the shortcoming that it was highly inefficient and impractical because in order to establish a definite relationship between audio data and image data, all the tracks on a still video floppy serving as a recording medium had to be searched to play all the audio tracks until an image track to be looked up was found, which inevitably took an enormous amount of time.

[0004]

② In addition, because only one image could be looked up from an audio track, when multiple images were involved, a single audio [track] could not be correlated with them. Therefore, there was the problem that an oral comment regarding a common concept could not be added to multiple images at one time.

③ Furthermore, because a large recording capacity was required to record audio data, there was a problem that, for cost reasons, it was especially inappropriate to use an audio track for a simple comment.

[0005]

Here, the problems pertaining to the aforementioned conventional digital electronic camera are also problems with a popular device that is used to process audio and image data. The present invention was thus made in view of the aforementioned problems, and its objective is to present a device for processing image data and audio data that is configured such that links can be established among plural image, audio, and text files to allow the respective files to be retrieved and reproduced mutually in order to allow expressions that are correlated with one another among the image, audio, and text files to be recorded and edited easily.

[0006]

Another objective is to present a device for processing image data and audio data, wherein sound or an image can be converted into a text format, so that a comment (identification information) can be added to the image using a small capacity. Yet another objective is to present a device for processing image data and audio data that allows the text that is added as a comment to be used for running a search and used as a keyword when generating a database.

[0007]

Furthermore, another objective is to present a device for processing image data and audio data by means of which such processing as automatic linking of data with information stored in another database can be realized.

[0008]

Means to solve the problems

In order to solve the aforementioned problems so as to achieve the objectives, the device for processing image data and audio data of the present invention is characterized in that it is equipped with a recording medium that records image data and audio data by correlating them, a specification means that specifies audio data stored in the aforementioned recording medium, a control means that retrieves image data that are correlated with the aforementioned specified audio data in the form of file information, and a reading means that reads the aforementioned retrieved image data and reproduces them.

[0009]

In addition, it is preferably characterized in that it is further equipped with a character recognition means that performs character recognition of the aforementioned image data, whereby it performs registration processing so as to record the result of recognition by said character recognition means in the form of a text file by correlating it with the aforementioned

respective image and audio data while allowing it to be retrieved. In addition, it is preferably characterized in that it is further equipped with a speech recognition means that recognizes the aforementioned audio data in the form of characters, whereby it performs registration processing so as to record the result of recognition by the aforementioned speech recognition means in the form of a text file by correlating it with the aforementioned respective image and audio data while allowing it to be retrieved.

[0010]

In addition, it is preferably characterized in that it is further equipped with a display means that displays the aforementioned respective data stored in the aforementioned recording medium in the form of file information, whereby mutually correlated information among the aforementioned respective files is edited, and the results of retrieval of the aforementioned respective data based on said correlated information are displayed.

[0011]

Function

Because the device for processing image data and audio data of the present invention is configured in the aforementioned manner, links can be established among more than one image, audio, and text file so as to allow the respective files to be retrieved and reproduced mutually, and expressions that are correlated with one another among the image, audio, and text files can be edited and recorded easily.

[0012]

In addition, because an audio or an image file can be converted into a text format, a comment (identification information) can be added to the image using a small capacity. In addition, the text that is added as a comment can be used for running a search and used as a keyword when generating a database. Furthermore, processing such as automatic data linking with information stored in another database can be realized.

[0013]

Application example

A preferred application example of the present invention will be explained in detail below with reference to attached figures. Figure 1 is a block diagram showing the configuration of a digital electronic camera of the present application example. In Figure 1, a memory card or a hard disk in compliance with the PCMCIA standard, for example, is used as recording medium 101. An audio jack and a speaker are used as audio input circuit 20 and audio output part 22,

respectively. A/D converter 24 is a device for converting an audio signal from an analog signal into a digital signal; and D/A converter 25 is a device for converting a digital audio signal sent from signal processing control CPU 13 into an analog signal. Switch 21 is a selection circuit for selecting when an audio signal is to be transmitted to audio output part 22. Memory bus controller 102 is used for transferring image data and audio data among pickup signal processing part 19, DSP 13 for signal processing control, image display buffer memory 12, and recording medium I/F circuit 104.

[0014]

D/A converter 26 is a device for converting digital image data from image display buffer memory 12 into an analog video signal, and video output part 23 is a video display device for displaying the analog-converted video signal as an image. 1 is a pickup lens, 2 is a diaphragm-shutter with a diaphragm function as well as a shutter function, 3 is a flash, mechanism-operation part control CPU 4 controls respective operating parts used for mechanical operations, and driver circuit 5 is a circuit for driving the respective mechanical components.

[0015]

Pickup element 6 is a CCD for converting light reflected from a photographic object into an electric signal, timing signal generator circuit 7 (hereinafter to be referred to as "TG") generates a timing signal that is needed to operate pickup element 6. Pickup element driver circuit 8 is a driver circuit for amplifying the signal from timing signal generator circuit 7 to the level at which an imaging signal can be driven, and preprocessing circuit 9 is equipped with a CDS circuit for removing output noise generated by pickup element 6 and a non-linear amplifier circuit to be operated before A/D conversion. A/D converter 10 converts the preprocessed data into a digital signal. CPU 13 for signal processing control is a signal processing control DSP (Digital Signal Processor) for controlling the signal processing part, operation display part 14 is a display part for providing indications to assist operations and show given camera statuses, and operation part 15 is an input device such as a keyboard that is used to control the camera from the outside. Recording medium I/F 104 is an I/F circuit used to connect the digital electronic camera based on the present application example with recording medium 101. Then a file format such as MS-DOS can be used as a recording file format with the recording medium of the camera of the present application example.

[0016]

Image display buffer memory 12 can be accessed from DSP 13 for signal processing control by pixel units, and any arbitrary operation means panel can be drawn while displaying a

pickup image. When a trackball is installed in operation part 15, a user I/F similar to the GUI (Graphical User Interface) of personal computers of recent years that utilizes a pointing device can be realized. That is, DSP 13 for signal processing control draws a variety of control panels in the image display buffer, and a user operates the pointing device (trackball in this case) on the control panels, whereby all types of camera operations can be performed. In addition, images, text, and a variety of diagrams are drawn at the image display part by DSP 13 for signal processing control in order to present the current camera status and all types of information, for example, management information regarding files in the recording medium, to the user. In the present application example, operations of the operation part 15 and aforementioned GUI-based user I/F will hereinafter be referred to as operations by operation part 1. In addition, a click or a double-click operation with the pointing device will be referred to as an activation operation.

[0017]

Monitoring of image recording in image recording mode

When DSP 13 for signal processing control detects a recording mode transition command that is entered by the user via operation part 15, DSP 13 for signal processing control executes the following processing in order to display a recording image on video output part 23. Mechanical operation part control CPU 4 and mechanical system driver circuit 5 are used to control the lens system as intended by the photographer. In such a case, photographing conditions are displayed on operation part 15 in order to inform the photographer of the camera status. Furthermore, a photometry circuit, not shown, is used to measure the brightness of the photographic object; and a value that indicates the opening of diaphragm-shutter 2 and shutter speed data is derived by the mechanical operation part control CPU 4. Diaphragm-shutter 2 is driven by mechanical system driver circuit 5 based on the control value derived by mechanical operation part control CPU 4. In addition, flash 3 may be activated, depending on the output of the not-shown photometry circuit. Exposure is performed as described above, and the light reflected from the object is incident on pickup element 6 via pickup lens 1 and diaphragm-shutter 2. In this case, diaphragm-shutter 2 is provided in order to regulate the amount of light incident on pickup element 6 and to prevent adverse effects of incident light on the signal charge while the image is being transferred when an interlaced read-type CCD is utilized as the pickup element. Pickup element 6 is operated based on a drive signal that is obtained by amplifying the output of TG 7 by pickup element driver circuit 8. Here, the operation of TG 7 is controlled by DSP 13 for signal processing control. The output of pickup element 6 driven in said manner is sent to preprocessing circuit 9. Preprocessing circuit 9 carries out CDS processing for removing low-frequency noise present in the output from the pickup element and processing for non-linearizing the pickup output in order to utilize the D range (digitized signal data) of A/D

converter 10 effectively. The preprocessed pickup signal output is converted to a digital signal at A/D converter 10, and it is converted into standard component video signals (for example, a brightness signal and two color difference signals or an RGB signal) before they are input to memory controller 102. Memory controller 102 transfers the digitized pickup signals to buffer memory 12 continuously under the control of DSP 13 for signal processing control.

[0018]

Figure 2 is a diagram showing a display format used in the image recording mode. When displaying a display image output to image display part 23, monitor region part 202 for the recorded image and parameter control panel part 203, where camera status and photographing conditions can be configured by the user, for example, are divided as shown in Figure 2. The digital data in image display buffer memory 12 are converted into an analog video signal by A/D converter 26 and output to video output part 23, and display is performed at video output part 23. The user can confirm the recorded image at monitor region part 202 in the form of a moving image.

[0019]

The moving image captured by pickup element 6 is displayed at video output part 23 through the aforementioned processing.

#### Recording of image

Once the camera is instructed to capture an image when operation part 15 is controlled by the photographer, DSP 13 for signal processing control stops displaying the moving image, accesses the image data held in the display buffer memory via the memory controller, applies digital compression processing, and records data on recording medium 101 via recording medium I/F 104.

[0020]

When memory bus controller 102 stops writing to image display buffer memory 12 for a prescribed interval of time after said transfer interval and recording is finished, the image to be recorded is displayed at video output part 23 as a still image. Thus, the user can confirm the still image that has just been recorded using video output part 23. In addition, once photographing is completed, to allow the content of the recorded image to be confirmed quickly, DSP 13 for signal processing control can add a thinned image, which is generated from the compressed image, to the compressed file. For example, the volume of the file does not increase very much



when an image is added that is thinned to approximately 1/8 of the original image dimensions in the horizontal and the vertical direction. Here, this image will be referred to as an index image.

[0021]

#### Monitoring and recording of audio during audio recording

While in the recording mode, DSP 13 for signal processing control can monitor audio to be recorded by connecting the output of audio input circuit 20 to the input of audio output part 22 using switch circuit 21. When DSP 13 for signal processing control detects an audio recording command that is entered by the user via operation part 15, DSP 13 for signal processing control receives the data that have been converted into digital data by A/D converter circuit 24, and transfers them to recording medium I/F 104 via memory bus controller 102.

[0022]

When DSP 13 for signal processing control detects cancellation of the audio recording command that is entered by the user via operation part 15, or when a prescribed amount of time has passed, DSP 13 for signal processing control assumes the cancellation of audio recording, and DSP 13 for signal processing control ends audio recording.

#### Simultaneous recording of images and audio

When recording images and audio simultaneously, as was already explained under <Recording images>, DSP 13 for signal processing control holds the audio data received from A/D converter 24 temporarily in an internal buffer of DSP 13 for signal processing control while the image is being compressed and transferred to recording medium I/F 104; it transfers the audio data to recording medium I/F 104 when the image transfer is completed. In the case of the NTSC format, what is called a vertical blanking interval, a 1.4-millisecond interval where no video signal is present, is provided before and after the vertical sync interval that is part of the 16.7-millisecond interval for 1 field. When image data are to be transferred at the normal video rate, image data that are not processed [sic] into a pickup signal are transferred during the time interval (approximately 15 milliseconds) that excludes said vertical blanking interval. The data are transferred at a rate of approximately 10 MBytes/sec to this end.

[0023]

When 1 8-bit sample of audio data is obtained at 22 kHz, the volume of data required for 16.7 milliseconds is approximately 370 Bytes. When said data are to be transferred using the remaining 1.4 milliseconds, a transfer rate of approximately 260 KBytes/sec is required. Said

speeds can be easily achieved when a memory card in compliance with the PCMCIA standard, for example, is used as a recording medium.

[0024]

As explained above, because the audio data are transferred during the vertical blanking interval out of the scanning interval of 1 screen, and the image data are transferred during the image interval by means of time-division multiplexing, the camera of the present application example can record images and audio simultaneously while monitoring them at the same time. Here, because the audio data can be assumed to be information correlated with an image, image and audio are filed such that they can be looked up mutually based on the explanation of <Linking audio and text to an image file> to be described later.

[0025]

In addition, the audio data may also be transferred during the horizontal blanking interval of the video signal. For example, when sampling audio under the conditions of 44.1 kHz, stereo, and 8 bits, a total of 6 bytes or 4 bytes of audio data are transferred and recorded during 1 horizontal blanking interval (1H interval). At this time, if a medium with a recording mechanism such as that of an HDD is used as a recording medium, the audio data should be recorded in an interleaved fashion at every 1H interval. However, when a semiconductor memory is used, there is no need to interleave the image data and the audio data at every 1H interval, and interleaved recording at every 1V interval may be used in order to record the audio data before the image data in 1 field interval (1V interval).

[0026]

#### Recording of text data

In the case of the camera of the present application example, 5 images and sounds are recognized and converted into text data in order to achieve significant reduction of the recording volume for comments on images.

#### ① Recording by means of character recognition

When DSP 13 for signal processing control detects a character recognition recording mode transition command that is entered by the user via operation part 15, the camera carries out the same processing as that under <Monitoring of recording in image recording mode> in order to output the images for character recognition to the image display part. As images to be displayed at image display part 301 at this time, image 302 to be recognized, part 303 where the character recognition result is shown, and parameter control panel 304 where the user can

configure camera status and photographing conditions are displayed in the divided manner shown in Figure 3.

[0027]

Then the following processing is repeated while execution of character recognition is requested by the user via operation part 15. DSP 13 for signal processing control pauses writing to the image display buffer memory and carries out character recognition of said image. Once character recognition of the image in the buffer memory is completed, DSP 13 for signal processing control displays the recognition result in display area 303 shown in Figure 3.

[0028]

When a satisfactory recognition result is obtained, the user ends the character recognition execution command via operation part 15. At this point, the user requests that the recognized text be finalized via operation part 15. The camera records the finalized text on recording medium 101 via memory controller 102 and recording medium I/F 104. If the text data are not finalized, a recognized image is rewritten into the buffer memory, and the aforementioned operations are repeated.

[0029]

#### ② Recording by means of speech recognition

When DSP 13 for signal processing control detects a character recognition recording mode transition command that is entered by the user via operation part 15, the camera carries out the same processing as that under <Monitoring and recording of audio during audio recording> in order to monitor the audio. At this time, as images to be displayed at image display part 401, part 402 where the speech recognition result is displayed and parameter control panel 403 where the user can configure camera status and photographing conditions are displayed in the divided manner shown in Figure 4. Furthermore, panel 130 may be configured in the manner shown in Figure 13 in a preferred specific example.

[0030]

Then, as instructed to execute the speech recognition by the user via operation part 15, DSP 13 for signal processing control receives the digital data converted by A/D converter 24, carries out the speech recognition processing, and displays the recognition result at display part 402 shown in Figure 4. The user ends the instruction for executing speech recognition via operation part 15 when the speech recognition result is obtained.

[0031]

If the recognition result is unsatisfactory, speech recognition is instructed again via operation part 15. When a satisfactory recognition result is obtained, the user gives an instruction to finalize the recognized text via operation part 15. The camera records the finalized text on the recording medium via memory controller 102 and recording medium I/F 104.

[0032]

As is clear from the explanation given above, the digital electronic camera of the present application example is capable of storing 3 kinds of files, for image data, audio data, and text data, on the recording medium.

Displaying files in the form of a lookup table

Figure 5 is a diagram showing a display format used to display a list of recorded files. When 3 kinds of files are to be displayed on the display part of the camera in the form of a video directory, DSP 13 for signal processing control reads the files recorded on the recording medium and displays index image 501 that represents an image file for images, and icons of the kinds indicated by 502 and 504 for audio and text, respectively, according to the types of file data, as shown in Figure 5. These indications may be displayed in the order they are recorded; or, only image, only audio, or only text may be displayed. In addition, when the user clicks on Erase button 509 after having selected an icon representing image, audio, or text using a pointing device, the camera erases the selected file.

[0033]

The respective icons 511 and 512 for audio and text that are provided below index image 501 are used to indicate whether text and/or audio that is linked to the image is present or not. Text and audio are linked according to <Linking of audio and text to image file>, to be described later. For example, the presence of data is indicated by showing icons 511 and 512 using a gray level. Said icons are selected using the pointing device in order to activate <audio playback> or <text playback>, to be described later.

[0034]

In Figure 5, portion 510 of the screen is formed into a scrollbar that is used for vertical scrolling in order to show objects such as index images and icons when they cannot all be shown on the screen. It functions in the same manner as a window display tool that has been utilized for personal computers and workstations in recent years.

### Image enlargement and playback

Figure 7 is a diagram showing a display format used for an enlarged image. In Figure 7, when a single image is to be enlarged for playback, the user selects a single index image using the pointing device provided in the operation part and gives an enlargement command (for example, by double-clicking a button). Upon detecting the aforementioned operation, DSP 13 for signal processing control reads the compressed image data from the selected image file, and expands and transfers these to the image display buffer memory in order to display them on the image display part. At this time, display part 701 for the enlarged image, respective control buttons 702-704, and respective icons 705 and 706 for multiple linked sounds and texts are displayed as shown in Figure 7.

[0035]

When button 702 is activated using the pointing device, the image panel displayed in Figure 7 is closed, and the aforementioned display shown in Figure 5 resumes.

### Audio playback

Next, in Figure 5, when audio is to be played back, the user selects and activates an Audio icon using the pointing device of the operation part.

[0036]

DSP 13 for signal processing control controls switch circuit 21 to connect the output from D/A converter 25 to the input of audio output part 22. DSP 13 for signal processing control controls recording medium I/F 104 and memory bus controller 102 so as to read the audio data and outputs them to the D/A converter at a sampling cycle for recording, and the audio is output from audio output part 22.

[0037]

### Text playback

Furthermore, when text is to be displayed, the user selects and activates a Text icon using the pointing device of the operation part. Figure 7 is a diagram showing a display format used for text data. DSP 13 for signal processing control controls recording medium I/F 104 and memory buffer controller 102 so as to read the text data, unfolds then into a character bit pattern, and transfers the text data of the kind shown in Figure 6 to the image display buffer memory in order to display them on the image display part.

[0038]

Text display part 601 and control button 602, among others, are displayed in Figure 6. When button 602 is activated using the pointing device, the panel in Figure 6 is closed.

#### Linking of audio and text to image files

The camera based on the present application example is equipped with control buttons indicated by icons 703 and 704 shown in Figure 7 that are used to add audio and text obtained through speech recognition, respectively, to the image when the single image is played back according to the explanation given in said <Enlargement and playback of independent images>. When icon 703 or 704 in Figure 7 is activated by clicking on it using the pointing device, DSP 13 for signal processing control records the audio or text in the same manner as that in <Monitoring and recording of audio during audio recording> or in <Recording by means of character recognition>. This operation can be realized by enlarging, playing, and monitoring the image data. Figure 13 is a diagram showing the condition when speech recognition control panel 130 is superimposed over the image display panel shown in Figure 7. For example, speech recognition control panel 130 can be displayed while it is superimposed over the image display panel shown in Figure 7 as shown in Figure 13. When recording image data and audio data simultaneously, or when recording audio and text while monitoring the playback result of the image data in the aforementioned manner, the image file and data to be looked up by the audio and text files are stored in [the same] file in order to create a mutually linked file. Figure 8 shows the respective data structures of image file and the audio and the text files within the file.

[0039]

In Figure 8, each file stores the numbers of linked files per each file format to be linked (the number of audio files is  $n$ , the number of text files is  $m$ , and the number of image files is  $k$ ) and file identifiers corresponding in number to the number of files. When this is done, because not only the audio files but also the image files look up the audio files and the text files, correlated audio and text files can be specified from a single image file in order to play or display them without the need to look up all the files, as was the case in the past.

[0040]

Where audio and text files linked to a single image file are present, when said image is to be displayed, the linked audio and text files can be displayed in the form of icons (for example, at the bottom of image data) as shown in Figures 5 and 7. When the user selects and activates said icons using the pointing device, the camera displays the correlated audio and text.

[0041]

Text is displayed in the manner shown in Figure 6. It is displayed independently, or it is superimposed over the display shown in Figure 7. In addition, when the user clicks on Delete button 707 after having selected icon 705 or 706 in Figure 7 using the pointing device, the camera can delete the correlated audio or text file. This operation means erases the information concerning the links established among the image file and the audio and text files shown in Figure 8. At this time, the audio and text files whose links are deleted may be present independently, or they may be erased.

[0042]

In addition, for independent display, the files are expressed in the form of a file lookup table comprising icons 502 and 504, for example, as shown in Figure 5.

Grouping of multiple image files and linking of audio and text files to the group

When multiple images are selected using the pointing device under said <Displaying of files in the form of a lookup table>, the index images are displayed with a bold frame as shown in Figure 9 in order to emphasize the fact that they have been selected. Furthermore, when Group button 901 in Figure 9 is clicked using the pointing device, a group of image files is created. At this time, the frames of the index images may be shown in a color different from that used for the other images in order to emphasize the fact that they have been grouped together.

[0043]

Furthermore, when Add audio button 902 or Add speech text button 903 shown in Figure 9 is activated using the pointing device, DSP 13 for signal processing control records the audio and/or the text in the same manner as that in the explanation given under <Monitoring and recording of audio during audio recording> or <② Recording by means of speech recognition>. At this point, multiple sounds and texts are added to the group. Comments on items unique to the group can be provided by means of these operations.

[0044]

Multiple images are linked to audio and text through the operations explained above. At this time, if reference data regarding all the other files is added to each file within the group, when this group is later specified, all the files have to be looked up, making management difficult. Thus, in the present application example, a group file for holding group information is generated when the generation of a group has become clear. The configuration of this file is shown in Figure 10. That is, it stores the number of the image, audio and text files that belong to

the group, and identifiers for the respective files. In addition, the image, audio and text files that belong to the group are organized in the manner shown in Figure 11 in order to allow looking-up the groups to which they belong. In other words, the individual files provide look-up for the group only, but they do not provide look-up for the files that belong to the group.

[0045]

When this kind of file configuration is adopted, group information, that is, files that belong to a group, can be identified quickly from any group, image, audio, or text file. For example, when retrieving other files that belong to the group from a single image file, once the group file to which it belongs is obtained, the identifiers of said files can be obtained directly.

[0046]

As is clear from the aforementioned application example, an image file can belong to multiple groups. When a group file is generated through the aforementioned procedure, the camera displays icon 505 that represents the group file in the <Lookup table of files>, as shown in Figure 5. When said icon is double-clicked using the pointing device, the files that belong to the group are displayed as shown in Figure 12. In Figure 12, when an Audio icon or Text icon is clicked on using the pointing device, the aforementioned <Playback of audio> or <Display means [sic] of text> is executed, whereby comments on/explanation of items unique to this group can be shown.

[0047]

To delete a file that belongs to the group, after the file to be deleted is selected using the pointing device in Figure 12, button 121 is used to activate Delete. Audio and text files deleted from the group remain present independently. When they are to be present independently, they are shown as icons 501, 502, and 504 in the lookup table in Figure 5.

[0048]

Also, Erase button 122 may be provided in order to erase the selected file when the button is activated. Needless to say, when deleted or erased from the group, the link information among the image, speech, and text data files shown in Figure 10 and Figure 11 is erased.

[0049]

In addition, to add a file to an existing group, the file to be added and the icon of the group to which the addition is made are selected using the pointing device in Figure 5, and Group is activated. Multiple files outside of the group can be selected during these operations.



Image files, audio files, and text files are present independently in the recording medium. Some of them can also be grouped together. Multiple files are selected using the pointing device while in <Lookup table of files>. A group file, to which images, sounds, and texts belong, is generated when Group button 508 in Figure 5 is activated using the pointing device.

[0050]

Effects of the present application example

As explained above, in the present application example, a single image or multiple images can be linked to audio and text, and their linking relationship can be reproduced quickly, so that mutually complementary expression of the image, audio, and text can be edited easily.

[0051]

In addition, because speech or an image is converted into text, a comment can be added using a small amount of capacity. Also, a search can be run based on the text that is added as a comment, and it can be used as a keyword when generating a database. Furthermore, such processing as automatic data linking with information stored in another database can be realized.

[0052]

Here, the present invention can be applied to a revision/modification of the aforementioned application example as long as it does not depart from its gist. For example, it may be applied to a system that is configured with multiple pieces of equipment, or it may be applied to a device that is comprised of a single piece of equipment. Needless to say, it can also be applied where a program is supplied to a system or a device.

[0053]

Effects of the invention

As explained above, because multiple image, audio, and text files can be linked together to allow the respective files to be retrieved and reproduced mutually so as to allow the respective files to be retrieved and reproduced mutually and quickly, the present invention offers the effect that mutually correlated expressions among the image, audio, and text files can be edited and recorded easily.

[0054]

Another effect is provided whereby because speech or an image can be converted into a text format, a comment (identification information) can be added using a small amount of capacity. Another effect is provided whereby a search can be run based on the text that is added

as a comment, and it can be used as a keyword when generating a database. Furthermore, yet another effect is provided whereby processing such as automatic data linking with information stored in another database can be realized.

### Brief description of the figures

Figure 1 is a block diagram showing the overall configuration of a digital electronic camera as an application example of the present invention.

Figure 2 is a diagram showing a display format used in an image recording mode.

Figure 3 is a diagram showing a display format used in a character recognition mode.

Figure 4 is a diagram showing a display format used in a speech recognition mode.

Figure 5 is a diagram showing a lookup table format of recorded files.

Figure 6 is a diagram showing a text display format.

Figure 7 is a diagram showing an image display format.

Figure 8 is a diagram showing internal file organizations when an image file and an audio file and a text file are mutually linked.

Figure 9 is a diagram showing a display format used when multiple files are selected in the displayed lookup table.

Figure 10 shows the data arrangement inside a group file.

Figure 11 shows the data arrangement inside each of the image, audio, and text files.

Figure 12 is a diagram showing a lookup table display format used for files that belong to a group.

Figure 13 is a diagram showing a display format used when a speech recognition control panel is superimposed over the image display format.

### Explanation of symbols

- 1 Pickup lens
- 6 Pickup element
- 10, 24 A/D converter
- 12 Buffer memory
- 13 DSP signal processing control
- 14 Operation display part
- 15 Operation part
- 19 Pickup signal processing part
- 20 Audio input circuit
- 22 Audio output part
- 23 Video output part

- 25, 26 D/A converter  
 101 Recording medium  
 102 Memory bus controller  
 104 Recording medium I/F

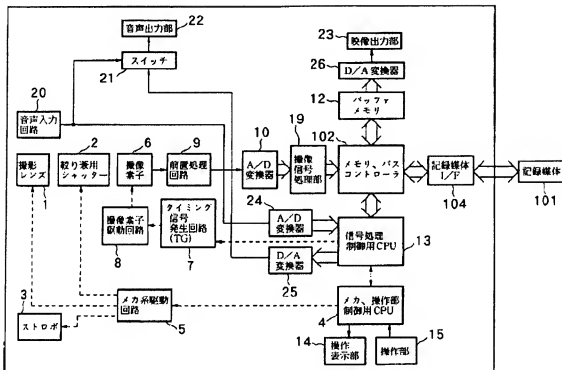


Figure 1

- Key:
- 1 Pickup lens
  - 2 Diaphragm-shutter
  - 3 Flash
  - 4 CPU for mechanical operation part control
  - 5 Mechanical system driver circuit
  - 6 Pickup element
  - 7 Timing signal generator circuit (TG)
  - 8 Pickup element driver circuit
  - 9 Preprocessing circuit
  - 10, 24 A/D converter
  - 12 Buffer memory
  - 13 CPU for signal processing control
  - 14 Operation display part
  - 15 Operation part
  - 19 Pickup signal processing part

- 20 Audio input circuit
- 21 Switch
- 22 Audio output part
- 23 Video output part
- 25, 26 D/A converter
- 101 Recording medium
- 102 Memory bus controller
- 104 Recording medium I/F

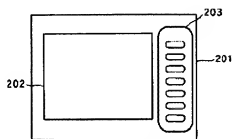


Figure 2

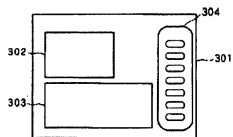


Figure 3

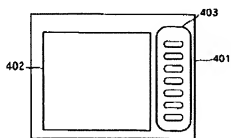


Figure 4

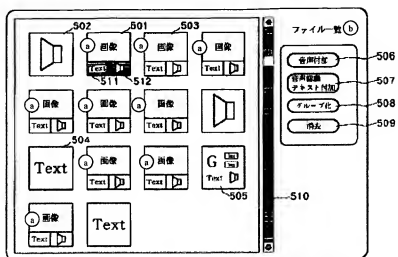


Figure 5

- Key: a Image  
 b List of files  
 506 Add audio  
 507 Speech recognition  
 508 Add text  
 509 Group  
 509 Erase

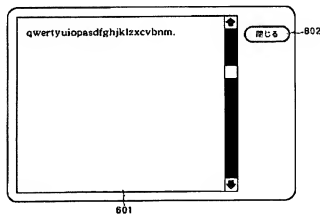


Figure 6

- Key: 602 Close

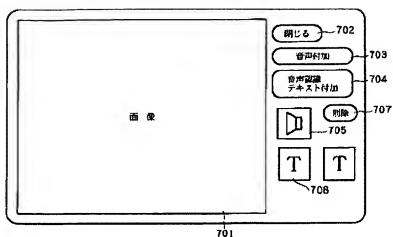


Figure 7

- Key:
- 701 Image
  - 702 Close
  - 703 Add audio
  - 704 Speech recognition Add text
  - 707 Delete

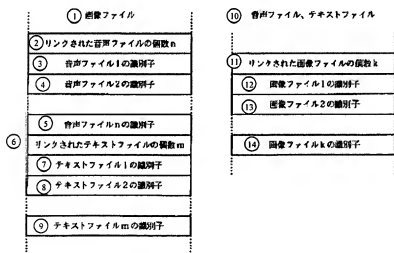


Figure 8

- Key:
- 1 Image file
  - 2 Number n of audio files linked
  - 3 Identifier of audio file 1
  - 4 Identifier of audio file 2
  - 5 Identifier of audio file n

- 6 Number m of text files linked
- 7 Identifier of text file 1
- 8 Identifier of text file 2
- 9 Identifier of text file m
- 10 Audio file, text file
- 11 Number k of image files linked
- 12 Identifier of image file 1
- 13 Identifier of image file 2
- 14 Identifier of image file k

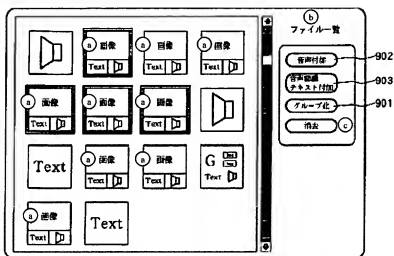


Figure 9

- Key:
- a Image
  - b List of files
  - c Erase
  - 901 Group
  - 902 Add audio
  - 903 Speech recognition
  - Add text

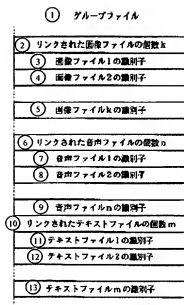


Figure 10

- Key:
- 1 Group file
  - 2 Number  $k$  of image files linked
  - 3 Identifier of image file 1
  - 4 Identifier of image file 2
  - 5 Identifier of image file  $k$
  - 6 Number  $n$  of audio files linked
  - 7 Identifier of audio file 1
  - 8 Identifier of audio file 2
  - 9 Identifier of audio file  $n$
  - 10 Number  $m$  of text files linked
  - 11 Identifier of text file 1
  - 12 Identifier of text file 2
  - 13 Identifier of text file  $m$



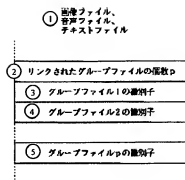


Figure 11

- Key: 1 Image file  
 Audio file  
 Text file  
 2 Number p of group files linked  
 3 Identifier of group file 1  
 4 Identifier of group file 2  
 5 Identifier of group file p

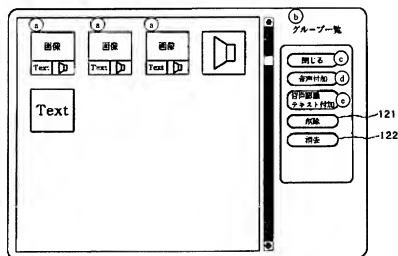


Figure 12

- Key: a Image  
 b List of groups  
 c Close  
 d Add audio  
 e Speech recognition  
 Add text  
 121 Delete

122 Erase

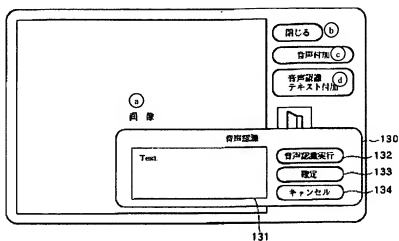


Figure 13

- Key: a Image  
 b Close  
 c Add audio  
 d Speech recognition  
 Add text  
 130 Speech recognition  
 132 Execute speech recognition  
 133 Finalize  
 134 Cancel



## 【特許請求の範囲】

【請求項1】 画像データと音声データとを関連づけて記録するための記録媒体と、前記記録媒体に格納された音声データを指定する指定手段と、

前記指定された音声データに関連する画像データをファイル情報として検索する制御手段と、前記検索された画像データを読み出し、再生する読出手段とを具備することを特徴とする画像データ及び音声データを処理する装置。

【請求項2】 前記画像データから文字認識を行なう文字認識手段を更に具備し、該文字認識手段による認識結果をテキストファイルとして前記画像及び音声の各データとに関連づけて記録し、かつ検索できるように登録処理を行うことを特徴とする請求項1に記載の画像データ及び音声データを処理する装置。

【請求項3】 前記音声データを文字として認識する音声認識処理手段を更に具備し、該音声認識処理手段による認識結果をテキストファイルとして前記画像及び音声の各データとに関連づけて記録し、かつ検索できるように登録処理を行うことを特徴とする請求項1に記載の画像データ及び音声データを処理する装置。

【請求項4】 前記記録媒体に格納された前記各データをファイル情報として表示する表示手段を更に具備し、前記各ファイル間の相互関連情報の編集、及び該関連情報に基づく前記各データの検索結果を表示することを特徴とする請求項1に記載の画像データ及び音声データを処理する装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は画像及び音声の処理して、記録再生を行なう装置に関するものである。

## 【0002】

【従来の技術】従来、画像及び音声データを処理し、記録する規格としてスチルビデオフォーマット方式が知られている。この従来のスチルビデオフォーマットは、画像データと音声データは夫々別のトラックにFM記録される。また、音声トラックは関連する画像トラックのトラック番号を記録するためのフィールドをコントロールコードに設けられており、特定の画像についての注釈（識別情報）を記録できるようにしたものである。

## 【0003】

【発明が解決しようとする問題点】しかしながら、上記のような従来のデジタル電子カメラにおいては、以下のような問題がある。

即ち、スチルビデオフォーマット方式では、音声トラックからの参照した画像トラックの指定はできるが、画像トラックからは関連する音声トラックを指定できない。従って、音声データと画像データとの関連を確定するために記録媒体であるスチルビデオフロッピー上の

全てのトラックを検索して全ての音声トラックを再生し、参照したい画像トラックを見つけ出さなければならざるに多大な時間を必要とするため無駄が多く現実的ではないという欠点があった。

【0004】また、音声トラックからは一つの画像しか参照できないので、複数の画像である場合に一つの音声に対応させることができなかった。従って、複数の画像に対して、共通の概念を音声で注釈しようとしても一度にできないという欠点があった。

更に、音声データを記録するためには多大な記録容量を必要とするので、特に単純な注釈を音声トラックに行うことはコスト的にみて不適切であるという問題があった。

【0005】尚、上述した従来のデジタル電子カメラに関する問題点は、一般的な音声及び画像データを処理する装置に係わる問題点でもある。従って、本発明は上述した課題に鑑みてなされたものであり、その目的とするところは、少なくとも1つ以上の画像、音声、テキストの各ファイル間のリンクを可能とし、それぞれのファイルを双方から高速に検索、再生できるように構成して、画像、音声、テキストファイル間での相互に関連性のある表現を容易に編集、記録できる音声及び画像データを処理する装置を提供することである。

【0006】また、音声あるいは画像をテキスト形式に変換できるため少ない容量で画像に注釈（識別情報）を付与することができる音声及び画像データを処理する装置を提供することである。また、注釈として付与されたテキストに基づいて検索を行ったり、データベースを作成するときのキーワードとして使用することも可能となる音声及び画像データを処理する装置を提供することである。

【0007】更に、他のデータベースに格納された情報との間で自動的にデータをリンクするなどという処理も可能となる音声及び画像データを処理する装置を提供することである。

## 【0008】

【課題を解決するための手段】上述の課題を解決し、目的を達成するために、本発明の画像データ及び音声データを処理する装置は、画像データと音声データとを関連づけて記録するための記録媒体と、前記記録媒体に格納された音声データを指定する指定手段と、前記指定された音声データに関連する画像データをファイル情報として検索する制御手段と、前記検索された画像データを読み出し、再生する読出手段とを具備することを特徴としている。

【0009】また、好ましくは前記画像データから文字認識を行なう文字認識手段を更に具備し、該文字認識手段による認識結果をテキストファイルとして前記画像及び音声の各データとに関連づけて記録し、かつ検索できるように登録処理を行うことを特徴としている。また、

好ましくは前記音声データを文字として認識する音声認識処理手段を更に具備し、該音声認識処理手段による認識結果をテキストファイルとして前記画像及び音声の各データとに関連付けて記録し、かつ検索できるように登録処理を行うことを特徴としている。

【0010】また、好ましくは前記記録媒体に格納された前記各データをファイル情報として表示する表示手段を更に具備し、前記各ファイル間での相互関連情報の編集、及び該関連情報に基づく前記各データの検索結果を表示することを特徴としている。

【0011】

【作用】以上のように、この発明に係る画像データ及び音声データを処理する装置は構成されているので、少なくとも1つ以上の画像、音声、テキストの各ファイル間のリンクが可能となり、それぞれのファイルを双方から高速に検索、再生できるので、画像、音声、テキストファイル間での相互に関連性のある表現を容易に編集、記録できる。

【0012】また、音声あるいは画像をテキスト形式に変換できるため少ない容量で画像に注釈（識別情報）を付与することができる。また、注釈として付与されたテキストに基づいて検索を行ったり、データベースを作成するときのキーワードとして使用することも可能となる。更に、他のデータベースに格納された情報との間で自動的にデータをリンクするなどという処理も可能となる。

【0013】

【実施例】以下、本発明の好適な実施例について、添付図面を参照して詳細に説明する。図1は、本実施例のデジタル電子カメラ本体のシステム構成を示すブロック図である。図1において、記録媒体101は、例えばPCMCIA規格に準拠したメモリカードやハードディスクなどである。音声入力回路20及び音声出力部22は、例えばオーディオジャックあるいはスピーカである。A/D変換器24は、音声信号をアナログ信号からデジタル信号に変換する装置で、D/A変換器25は、信号処理制御用CPU13から送られるデジタルの音声信号をアナログに変換する装置である。スイッチ21は音声出力部22への音声信号の送信を選択する選択回路である。メモリバスコントローラ102は、撮像信号処理部19、信号処理制御用DSP13、画像表示用バッファメモリ12、記録媒体I/F回路104の間での画像データ、音声データなどの転送を行う。

【0014】D/A変換器26は画像表示用バッファメモリ12からのデジタル画像データをアナログ映像信号に変換する装置であり、映像出力部23はアナログに変換された映像信号を映像として表示する映像表示装置である。1は撮影レンズであり、2は絞り機能とシャッター機能を兼ねる絞り兼用シャッター、3はストロボ、メカ及び操作部制御用CPU4は機械的な操作を行う各

操作部を制御し、駆動回路5はメカ系各部の駆動させる回路である。

【0015】撮像素子6は被写体からの反射光を電気信号に変換するCCDで、タイミング信号発生回路7は撮像素子6を動作させるために必要なタイミング信号を発生する（以下、「TG」と称する）。撮像素子駆動回路8はタイミング信号発生回路7からの信号を撮像素子の駆動可能なレベルに増幅する駆動回路で、前置処理回路9は撮像素子6で発生する出力ノイズ除去のためのCCD回路及びA/D変換前に行う非線形増幅回路を備える。A/D変換器10は前置処理後のデータをデジタルに変換する。信号処理制御用CPU13は信号処理部を制御する信号処理制御用DSP（DIGITAL SIGNAL PROCESSOR）であり、操作表示部14は操作補助のための表示やカメラの状態を表す表示部で、操作部15はカメラを外部から制御するためのキーボード等の入力装置である。記録媒体I/F回路104は本実施例に基づくデジタル電子カメラと記録媒体101とを接続するための記録媒体I/F用回路である。そして、本実施例のカメラの記録媒体への記録ファイルフォーマットは、例えばMS-DOSのようなファイルフォーマットを使用することができる。

【0016】画像表示用バッファメモリ12は信号処理制御用DSP13から画素単位でアクセスすることが可能であり、また撮影する画像を表示しながら任意の操作手段パネルを描画することができる。操作部15にトラックボールなどを実装すれば、近年におけるパーソナルコンピュータにおけるポインティングデバイスを用いたGUI（Graphical User Interface）と同様のユーザーI/Fを実現できるように becoming。つまり信号処理制御用DSP13が画像表示用バッファに様々なコントロールパネルを描画して、ユーザーがそのコントロールパネルに対してポインティングデバイス（この場合トラックボール）によって操作することでカメラのあらゆる操作が可能となる。また信号処理制御用DSP13によって、画像やテキストや様々な図形を画像表示部に描画することでユーザーに対して撮影された画像、カメラの現在の状態、記録媒体上のファイル管理情報など多種多様な情報の提示が行われる。本実施例では以下操作部15と上記GUIによるユーザーI/F操作を操作部15による操作と称することとする。またポインティングデバイスによるクリックあるいはダブルクリック操作などを起動操作と称することとする。

【0017】<画像記録モード時の記録画像のモニタ> 信号処理制御用DSP13がユーザーの操作部15の記録モード移行命令を検知すると、信号処理制御用DSP13は以下の処理を実行し記録する画像を映像出力部23に表示する。撮影者の意図に応じてレンズ系の制御がメカ操作部制御用CPU4及びメカ系駆動回路5により行われる。この際、撮影条件などが操作部15に表示さ

れ、撮影者にカメラの状況を伝える。さらに不図示の測光回路により被写体の明るさを測定し、絞り兼用シャッター2の絞り度を表す値やシャッタースピードのデータをメカ操作部制御用CPU1にて導出する。メカ操作部制御用CPU4で導出された制御値に基づき、メカ系駆動回路5により絞り兼用シャッター2を駆動する。また測光回路(不図示)の出力によってはストロボ3を発光させて撮影することになる。このようにして露光されて、被写体の反射光が撮影レンズ1及び絞り兼用シャッター2を介して撮像素子6に入射される。この際、絞り兼用シャッター2は撮像素子6への入射光量を制限するとともに、撮像素子としてインターレース読み出し型CCDを用いた場合、画像の転送途中に入射光が信号電荷に悪影響を与えないようにするために設けられている。撮像素子6は、TG7からの出力を撮像素子駆動回路8によって増幅された駆動信号により動作させる。尚、TG7は信号処理制御用DSP13によりその動作を制御されている。以上のようにして駆動された撮像素子6の出力は前置処理回路9に送られる。前置処理回路9では撮像素子からの出力に発生する低域ノイズを除去するCDS処理と共に、A/D変換器10のDレンジ(デジタル化された信号データ)を有効に用いるために撮像素子6を非線形化する処理を行っている。前置処理された撮像素子出力はA/D変換器10においてデジタル信号に変換され、撮像素子処理部19によって標準的なコンポーネントビデオ(例えば輝度信号と2色差信号やRGB信号)に変換されメモリコントローラ102に入力される。メモリコントローラ102では信号処理制御用DSP13の制御により、デジタル化された撮像素子信号をバッファメモリ12に連続的に転送していく。

【0018】図2は、画像記録モードにおける表示形式を示す図である。画像表示部23に出力される表示画像は例えば図2に示すように記録画像のモニタ領域部分202と、カメラの状態や撮影条件などユーザーが設定できるパラメータのコントロールパネル部分203を分割されて表示される。画像表示用バッファメモリ12のデジタル化されたデータはA/D変換器26によってアナログ映像信号に変換され映像出力部23に出力され、映像出力部23で表示される。ユーザーは、記録する画像を動画像としてモニタ領域部分202において確認できる。

【0019】以上の処理によって撮像素子6によって捕らえられた動画像が映像出力部23に表示されるのである。

<画像の記録>撮影者が操作部15を制御することによりカメラに撮影を指示すると、信号処理制御用DSP13は動画像表示を停止して表示用バッファメモリ12に保持された画像データにメモリコントローラを介してアクセスしてデジタルの圧縮処理を行った後、記録媒体I/F104を介して記録媒体101に記録する。

【0020】この転送期間及び記録終了後一定期間中メモリバスコントローラ102は画像表示用バッファメモリ12への書き込みを停止すれば記録する画像が静止されて映像出力部23に表示される。従ってユーザーがたった今記録した静止画像を映像出力部23で確認できる。また撮影後、より速く記録された画像の中身を確認するために、信号処理制御用DSP13は圧縮した画像の間引き画像を圧縮ファイルに付加することができる。例えば、元の画像の縦横8分の1程度に間引いた画像を付加してもファイル容量はそれほど増加しないのである。尚、この画像を索引画像と呼ぶ。

【0021】<音声記録時の音声のモニタ及び記録>記録モード時には信号処理制御用DSP13はスイッチ回路21で音声入力回路20の出力を音声出力部22の入力に接続して記録する音声をモニタすることができ、信号処理制御用DSP13がユーザーの操作部15による音声記録命令を検知すると、信号処理制御用DSP13はA/D変換器24でデジタルデータに変換したデータを受け取り、メモリバスコントローラ102を介して記録媒体I/F104に転送する。

【0022】信号処理制御用DSP13がユーザーの操作部15による音声記録命令の解除を検知するか、あるいは信号処理制御用DSP13が一定時間経過したことで音声記録の解除とみなした時点で信号処理制御用DSP13は音声の記録を終了する。

<画像と音声の同時記録>画像と音声を同時に記録する場合、前述の<画像の記録>で説明したように画像を圧縮して記録媒体I/F104に転送している間も、信号処理制御用DSP13はA/D変換器24から受け取った音声のデータを信号処理制御用DSP13の内部バッファに一時的に保存しておき、画像の転送が終了した時点で音声データを記録媒体I/F104に転送する。NTSC方式の1フィールド期間16.7ミリ秒のうち、垂直同期期間とその前後には1.4ミリ秒程度の垂直ブランキング期間といわれる映像信号が存在しない期間がある。画像データを通常のビデオレートで転送する場合、この垂直ブランキング期間を除いた時間(約15ミリ秒)の間に撮像素子処理されている画像データを転送する。これを実現するために約10MByte/secのスピードでデータを転送する。

【0023】音声データを1サンプリング8ビットで22kHzでサンプリングするとき16.7ミリ秒分のデータ容量は約370Byteである。このデータを残りの1.4ミリ秒で転送する場合、約260KByte/secの転送スピードが必要である。これらの転送スピードは、例えば記録媒体としてPCMCI A規格などに準拠したメモリーカードならば十分満足できる速度である。

【0024】以上説明したように、1画面の走査期間のうち垂直ブランキング期間中に音声データを転送し、映像期間に画像転送を行い時分割多重することで本実施例

のカメラは画像と音声と同時にモニタしながら同時に記録できる。このときは音声データは画像に関連した情報とみなせるため、画像及び音声を後述する<画像ファイルへの音声及びテキストのリンク>の説明に基づいて、両者が相互に参照し合うようにファイル管理する。

【0025】また音声データを映像信号の水平ブランキング期間内に転送しても良い。例えば、4.4、1kHz、ステレオ、8ビットの条件で音声をサンプリングするとき、1回の水平ブランキング期間(1H期間)内に計6Byteもしくは4Byteの音声データを転送・記録する。この時、記録媒体としてHDDなどの記録機構を有する媒体を用いる際には1H期間毎に音声データがインターリーブされるように記録するのが良いが、半導体メモリの場合には画像データと音声データを1H毎にインターリーブさせる必要はなく、1フィールド期間(1V期間)の画像データの前に音声データが記録されるように1V毎のインターリーブ記録としても良い。

【0026】<テキストデータの記録>本実施例のカメラは5つの画像や音声を認識してテキストデータにすることで画像に対する注釈のための記録容量の拡大を大幅に削減する。

<文字認識による記録>信号処理制御用DSP13がユーザーの操作表示部15による文字認識記録モード移行命令を感知するとカメラは<画像記録モード時の記録モニタ>と同じ処理を行い文字認識する画像を画像表示部に出力する。このとき画像表示部301に表示される画像は、図3のように認識しようとする画像302、文字認識された結果を表示する部分303、カメラの状態や撮影条件などをユーザーが設定できるパラメータのコントロールパネル304などを分割して表示する。

【0027】そしてユーザーが操作部15によって文字認識実行を指示している間は以下の処理を繰り返す。信号処理制御用DSP13は一旦画像表示用バッファメモリへの書き込みを停止してその画像に対して文字認識処理を行う。バッファメモリ上の画像に対する文字認識処理が終了すると信号処理制御用DSP13はその認識結果を図3に示す表示領域303に表示する。

【0028】ユーザーは満足な認識結果が得られたところで操作部15によって文字認識実行の指示を終了する。この時点でユーザーは操作部15によって認識されたテキストの確定を指示する。カメラは確定されたテキストをメモリコントローラ102、記録媒体I/F104を介して記録媒体101に記録する。テキストデータが確定されない場合は、再度認識画像をバッファメモリに書き込み直し上述の動作を繰り返す。

【0029】<音声認識による記録>信号処理制御用DSP13がユーザーの操作部15による文字認識記録モード移行命令を感知するとカメラは<音声記録時の音声のモニタ及び記録>に説明したのと同じ処理を行い音声をモニタする。このとき画像表示部401の表示画像

は図4のように音声認識された結果を表示する部分402、カメラの状態や、認識条件等ユーザーが設定できるパラメータのコントロールパネル403などを分割して表示する。さらに好適な具体例として、図13のバネル130のように構成できる。

【0030】そしてユーザーが操作部15によって音声認識実行を指示している間信号処理制御用DSP13はA/D変換器24でデジタルデータに変換したデータを受け取り、音声認識処理を行い認識結果を図4の表示部402に表示する。ユーザーは音声の認識結果が得られたところで操作部15によって音声認識実行の指示を終了する。

【0031】認識結果に不満がある場合は再度操作部15によって音声認識実行を指示する。満足する認識結果が得られたらこの時点でユーザーは操作部15によって認識されたテキストの確定を指示する。カメラは確定されたテキストをメモリコントローラ102、記録媒体I/F104を介して記録媒体に記録する。

【0032】以上の説明により明らかなるように本実施例のデジタル電子カメラは記録媒体に画像データ、音声データ、テキストデータの3種類のファイルを保持できる。

<ファイルの一覧表示>図5は、記録されたファイルの一覧を表示する表示形式を示す図である。3種のファイルをカメラの表示部に映像としてディレクトリ表示するときは信号処理制御用DSP13が記録媒体上に記録されているファイルを読み出し、ファイルデータの種類に応じて図4に示すように画像については画像ファイルを表示する索引画像501、音声、テキストはそれぞれ502、504のようなアイコンとして表示する。これらの表示は記録時間順に表示したり、また画像だけ、音声だけ、テキストだけの表示を行うこともできる。またユーザーは索引画像、音声、テキストを表すアイコンをポインティングデバイスで選択したのち、消去ボタン509をクリックするとカメラは選択されたファイルの消去を行う。

【0033】索引画像501の下部にある音声、テキストの夫々のアイコン511、512はその画像ファイルにリンクされたテキストや音声が存在するかどうかを示すためのものである。テキスト及び音声は後述する<画像ファイルへの音声、テキストのリンク>に従ってリンクされる。例えば、511、512のようにアイコンをグレーレベルにすることでデータが存在することを表現する。これらのアイコンをポインティングデバイスを用いて選択して起動させ、後述する<音声の再生>又は<テキストの再生>が行われる。

【0034】図5において、画面の一部510は表示する索引画像やアイコンなどのオブジェクトが画面に入りきらないときにそれらを検索できるように上下にスクロールするためのスクロールバーである。これは近年のパ

ーソナルコンピュータやワークステーションで用いられるウィンドウ表示のためのツールと同様に機能する。

＜画像の拡大及び再生＞図7は、拡大された画像の表示形式を示す図である。図7において一つの画像について拡大再生したい場合、ユーザーは操作部のポインティングデバイスを用いて一つの索引画像を選択して拡大命令（例えばボタンのダブルクリック）を行う。信号処理制御用DSP13は上記操作を検知すると選択された画像ファイルから圧縮画像データを読み出し、伸長して画像表示用バッファメモリに転送し画像表示部に表示する。このときの表示は、例えば図7に示すように拡大された画像の表示部701と各制御用ボタン702～704、更にリンクされた複数の音声、テキストのアイコンとして夫々705、706などが表示される。

【0035】ポインティングデバイスボタン702を起動すると図7に表示された画像のパネルは閉じられて前述の図5の表示に戻る。

＜音声の再生＞次に、図5において、音声を再生したい場合、ユーザーは操作部のポインティングデバイスを用いて音声アイコンを選択して起動する。

【0036】信号処理制御用DSP13はスイッチ回路21を制御してD/A変換器25からの出力を音声出力部22の入力に接続する。信号処理制御用DSP13は記録媒体1/F104、メモリバスコントローラ102を制御して音声データを読み出しD/A変換器に記録時の標本化周期で出力し音声出力部22から音声が出力される。

【0037】＜テキストの再生＞更に、テキストを表示したい場合、ユーザーは操作部のポインティングデバイスを用いてテキストアイコンを選択して起動する。図7は、テキストデータの表示形式を示す図である。信号処理制御用DSP13は記録媒体1/F104、メモリバスコントローラ102を制御してテキストデータを読み出しキャラクタビットパターン展開して例えば図6に示すようなテキストデータを画像表示用バッファメモリに転送し画像表示部に表示する。

【0038】図6ではテキストの表示部601と制御用ボタン602などが表示される。ポインティングデバイスでボタン602を起動すると第6図のパネルは閉じられる。

＜画像ファイルへの音声及びテキストのリンク＞本実施例に基づくカメラは前述の＜単独の画像の拡大及び再生＞での説明に従って一つの画像を再生したとき、画像に対して音声、あるいは音声認識によるテキストの付加を行うためのコントロールボタンを図7に示すアイコン703、704を夫々備えている。ポインティングデバイスで図7中のアイコン703、704をクリックし起動すると信号処理制御用DSP13は夫々＜音声記録時の音声のモニタ及び記録＞、＜音声認識による記録＞と同様に音声、テキストの記録を行う。この操作は画像デ

ータを拡大及び再生しモニタすることによって行うことができる。図13は、図7の画像表示パネルの上に音声認識の制御用パネル130を重ねて表示した様子を示す図である。例えば、図13のように図7の画像表示パネルの上に音声認識の制御用パネル130を重ねて表示することができる。画像データと音声データを同時記録する時、及び上記のように画像データを再生した結果をモニタしながら音声、テキストを記録する時は画像ファイル及び音声、テキストファイル双方がお互いを参照するデータをファイル内に格納することによって相互にリンクされたファイルとなる。図8は、ファイル内の画像ファイル及び音声ファイルとテキストファイル夫々のデータ構成を示している。

【0039】図8において、夫々のファイルはリンクされるファイル形式ごとにリンクされたファイル数（夫々音声ファイル数をn、テキストファイル数をm、画像ファイル数をk）及びファイル数の応じたファイルの識別子を格納する。このようにすれば音声ファイルだけでなく画像ファイルも音声、テキストファイルを参照しているので従来のように全てのファイルを検索しなくても一つの画像ファイルから関連される音声、テキストファイル特定してそのファイルを再生あるいは表示することができる。

【0040】一つの画像ファイルにリンクされた音声、テキストが存在する場合、その画像を表示する場合夫々図5、図7に示すようにリンクされた音声ファイル、テキストファイルをアイコンで（例えば、画像データの下部に）表示することができる。ユーザーはこのアイコンをポインティングデバイスで選択、起動するとカメラは関連づけられた音声の再生、テキストの表示を行う。

【0041】テキストは図6のように表示される。この表示は単独、または図5、図7の表示の上に重ねて表示される。またユーザーは図7においてアイコン705、706をポインティングデバイスで選択した後、削除ボタン707をクリックするとカメラは関連づけられた音声、テキストファイルの削除を行うことが可能である。この操作手段は図8に示した画像ファイル、音声、テキストファイルのお互いのリンク情報を消去する。このときリンクの削除された音声やテキストファイルは単独で存在するようにしてもよいし、消去するようにしてもよい。

【0042】また、単独で表示させるには、図5においてアイコン502、504のファイル一覧表示のように夫々表現される。

＜複数の画像ファイルのグループ化、及びグループへの音声、テキストのリンク＞前述の＜ファイルの一覧表示＞において、ポインティングデバイスによって複数の画像を選択すると索引画像は選択されたことを強調するように例えば図9のように枠が夫々表示される。更に、ポインティングデバイスで図9のグループ化ボタン901



をクリックすると画像ファイルのグループが作成される。この時グループ化されたことを強調するために索引画像の枠の色を他の画像と異なる色にすることもできる。

【0043】更に、ポインティングデバイスで図9に示す音声付加ボタン902あるいは音声テキスト付加ボタン903を起動すると信号処理制御用DSP13は夫々＜音声記録時の音声のモニタ及び記録＞、＜音声認識による記録＞での説明と同様に音声、テキストの記録を行う。この時点で複数の音声、テキストがグループに付加されることとなる。この操作によってグループに特有な事柄についての注釈説明を与えることができる。

【0044】以上説明した操作により複数の画像と音声、テキストがリンクされる。この時グループ内の夫々のファイルについてそれ以外の全てのファイルに対する参照データを付加すると、のちにこのグループを特定するために全てのファイルを検索しなければならず管理が複雑になる。そこで本実施例ではグループの発生が明確になった時点でグループ情報を保持するためのグループファイルを作成する。このファイルの構成は、図10に示すようになる。即ち、グループに属する画像、音声、テキスト夫々のファイルの個数、及び夫々のファイルの識別子が格納される。またグループに属する画像、音声、テキストファイルは夫々のファイルが属するグループへの参照ができるように図11のように編成される。即ち、個々のファイルはグループを参照するだけでそのグループに属するファイルへの参照は行わない。

【0045】このようなグループのファイル構成により、グループ、画像、音声、テキスト、どのファイルからでもグループ情報をなわちグループに属するファイルを高速に同定できる。例えば一つの画像ファイルからそのグループに属する他のファイルを探する場合、一旦所属するグループファイルを獲得してしまえばそれらのファイルの識別子を直接獲得できる。

【0046】以上説明した実施例から明らかなように画像ファイルは複数のグループに属することが可能である。上述の手順によってグループファイルが発生されるとカメラは＜ファイルの一覧表＞において、図5のようにグループファイルを表示するアイコン505を表示する。ポインティングデバイスでこのアイコンをダブルクリックすると図12のようにグループに属するファイルが表示される。図12において、音声アイコン、テキストアイコンをポインティングデバイスでクリックして起動すれば夫々前述した通りの＜音声の再生＞、＜テキストの表示手段＞が行われ、このグループに特有な事柄についての注釈、説明として表現することができる。

【0047】グループに属するファイルをグループから削除するためには図12において削除するファイルをポインティングデバイスで選択した後、ボタン121で削除を起動する。グループから削除された音声やテキスト

ファイルは単独で存在するようにする。単独で存在させるようにするときは図5においてアイコン501、502、504のように一覧表示に現れるようになる。

【0048】また消去ボタン122を設定してこのボタンを起動することで選択されたファイルを消去してもよい。グループから削除、あるいは消去される場合はもちろん、図10、図11に示されるグループファイル、画像、音声テキストデータファイルとの間のリンク情報は消去される。

【0049】またファイルを既存のグループに加えるためには図5において加えるファイル及び加える先のグループアイコンをポインティングデバイスで選択してグループ化を起動する。これらの操作においてグループ以外のファイルは複数選択することが可能である。記録媒体には単独で画像ファイル、音声ファイル、テキストファイルが存在する。これらのいくつかをグループ化することも可能である。＜ファイルの一覧表＞においてポインティングデバイスによって複数のファイルを選択する。さらにポインティングデバイスで第5図のグループボタン508を起動することで画像、音声、テキストが属するグループファイルを作成する。

【0050】（本実施例の効果）以上説明したように、本実施例においては、単一の画像又は複数の画像と、音声、テキスト間のリンクが可能になりそのリンクされた関係を高速に再現できるので、画像、音声、テキストの相互補間的な表現を容易に編集することができる。

【0051】また音声あるいは画像をテキストに変換するため少ない容量で画像に注釈を付けることができる。また注釈としてつけられたテキストをもとに検索を行ったり、データベースを形成するときのキーワードとすることも可能となる。さらには他のデータベースの情報と自動的にリンクするなどの処理も可能となる。

【0052】尚、本発明は、その趣旨を逸脱しない範囲で上記実施例を修正又は変形したものに適用可能である。例えば、複数の機器から構成されるシステムに適用しても1つの機器から成る装置に適用してもよい。また、システム或は装置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

#### 【0053】

【発明の効果】以上説明したように、本発明によれば、少なくとも1つ以上の画像、音声、テキストの各ファイル間のリンクが可能となり、それぞれファイルと双方から高速に検索、再生できるので、画像、音声、テキストファイル間での相互に関連性のある表現を容易に編集、記録することができ効果がある。

【0054】また、音声あるいは画像をテキスト形式に変換できるため少ない容量で画像に注釈（識別情報）を付与することができる効果がある。また、注釈として付与されたテキストに基づいて検索を行ったり、データバ

ースを作成するときのキーワードとして使用することも可能となる効果がある。更に、他のデータベースに格納された情報との間で自動的にデータをリンクするなどという処理も可能となる効果がある。

【図面の簡単な説明】

【図1】本発明の実施例のデジタル電子カメラの全体構成を示すブロック図である。

【図2】画像記録モードにおける表示形式を示す図である。

【図3】文字認識モードにおける表示形式を示す図である。

【図4】音声認識モードにおける表示形式を示す図である。

【図5】記録されたファイルの一覧表示形式を示す図である。

【図6】テキストの表示形式を示す図である。

【図7】画像の表示形式を示す図である。

【図8】画像ファイルと、音声ファイル及びテキストファイルとが相互にリンクされた場合の双方のファイル内部の編成を示す図である。

【図9】一覧表示において複数のファイルが選択された場合の表示形式を示す図である。

【図10】グループファイルの内部データ編成である。

【図11】グループに属する画像、音声、テキスト各ファイルの内部のデータ編成を示す図である。

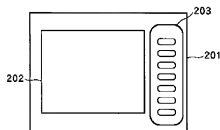
【図12】グループに属するファイルの一覧表示形式を示す図である。

【図13】画像表示形式の上に音声認識制御パネルを重ねて表示したときの表示形式を示す図である。

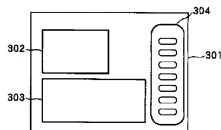
【符号の説明】

- 1 撮影レンズ
- 6 撮像素子
- 10, 24 A/D変換器
- 12 バッファメモリ
- 13 信号処理制御用DSP
- 14 操作表示部
- 15 操作部
- 19 撮像信号処理部
- 20 音声入力回路
- 22 音声出力部
- 23 映像出力部
- 25, 26 D/A変換器
- 101 記録媒体
- 102 メモリバスコントローラ
- 104 記録媒体I/F

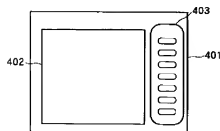
【図2】



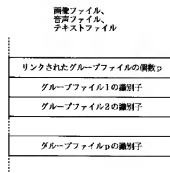
【図3】



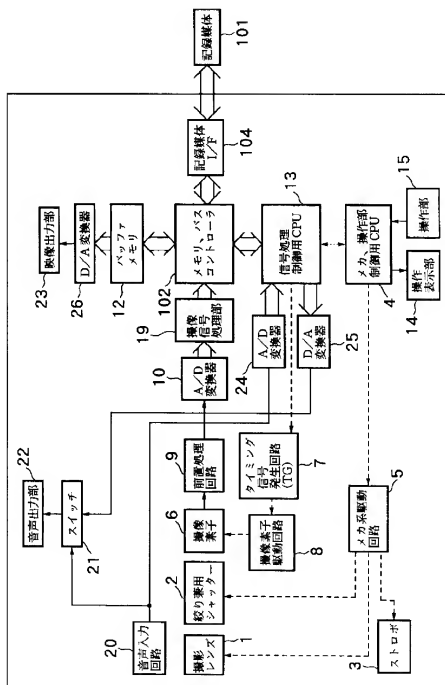
【図4】



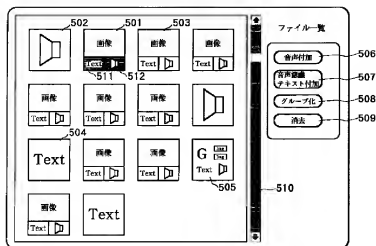
【図11】



【图 1】



【図5】

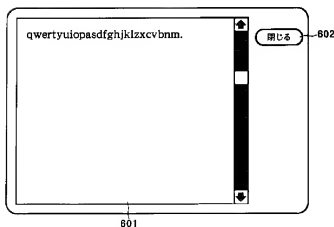


【図10】

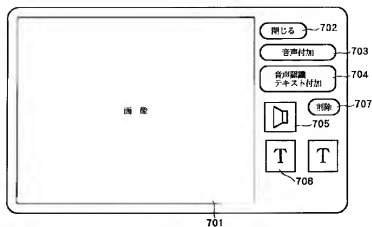
グループファイル

リンクされた画像ファイルの個数 k
画像ファイル1の識別子
画像ファイル2の識別子
画像ファイルkの識別子
リンクされた音声ファイルの個数 n
音声ファイル1の識別子
音声ファイル2の識別子
音声ファイルnの識別子
リンクされたテキストファイルの個数 m
テキストファイル1の識別子
テキストファイル2の識別子
テキストファイルmの識別子

【図6】



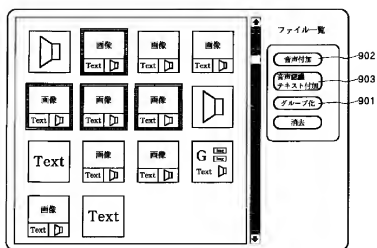
【図7】



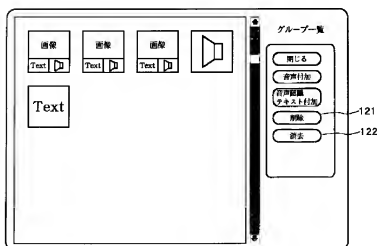
【図8】

画像ファイル	音声ファイル、テキストファイル
リンクされた音声ファイルの個数 $n$	リンクされた画像ファイルの個数 $k$
音声ファイル1の識別子	画像ファイル1の識別子
音声ファイル2の識別子	画像ファイル2の識別子
音声ファイル $n$ の識別子	
リンクされたテキストファイルの個数 $m$	画像ファイル $k$ の識別子
テキストファイル1の識別子	
テキストファイル2の識別子	
テキストファイル $m$ の識別子	

【図9】



【図12】



【図13】

